

Development and Application of an Exploratory Reduced Reparameterized Unified Model

Steven Andrew Culpepper

University of Illinois at Urbana-Champaign

Yinghan Chen

University of Nevada, Reno

Exploratory cognitive diagnosis models (CDMs) estimate the \mathbf{Q} matrix, which is a binary matrix that indicates the attributes needed for affirmative responses to each item. Estimation of \mathbf{Q} is an important next step for improving classifications and broadening application of CDMs. Prior research primarily focused on an exploratory version of the restrictive deterministic-input, noisy-and-gate model, and research is needed to develop exploratory methods for more flexible CDMs. We consider Bayesian methods for estimating an exploratory version of the more flexible reduced reparameterized unified model (rRUM). We show that estimating the rRUM \mathbf{Q} matrix is complicated by a confound between elements of \mathbf{Q} and the rRUM item parameters. A Bayesian framework is presented that accurately recovers \mathbf{Q} using a spike-slab prior for item parameters to select the required attributes for each item. We present Monte Carlo simulation studies, demonstrating the developed algorithm improves upon prior Bayesian methods for estimating the rRUM \mathbf{Q} matrix. We apply the developed method to the Examination for the Certificate of Proficiency in English data set. The results provide evidence of five attributes with a partially ordered attribute hierarchy.

Keywords: *exploratory cognitive diagnosis modeling; reduced reparameterized unified model; Bayesian; spike-slab priors*

Introduction

Cognitive diagnosis models (CDMs) offer a useful psychometric framework for classifying individuals on a collection of fine-grained binary attributes. A critical component for applying existing CDMs is the specification of a binary \mathbf{Q} matrix, which indicates the attributes required for each item. Correctly specifying \mathbf{Q} is essential for ensuring accurate diagnoses and model parameter estimates (Henson & Templin, 2007a; Rupp & Templin, 2008). Prior CDM research mainly relied upon content expert knowledge to specify \mathbf{Q} . However, specifying

\mathbf{Q} is a challenging task. Cognitive theory may be too underdeveloped to offer guidance for specifying \mathbf{Q} . Additionally, even in cases where cognitive theory is available, there is no guarantee that content experts will agree on all entries of a \mathbf{Q} matrix. The general unavailability of \mathbf{Q} for most content areas and data sets poses a barrier to widespread applications of CDMs.

Clearly, a fundamental problem for psychometric research and practice is the estimation of \mathbf{Q} . Recent research addressed this critical issue for CDMs by developing procedures for validating a specified \mathbf{Q} (Chiu, 2013; de la Torre, 2008; de la Torre & Chiu, 2016), estimating a subset of elements of \mathbf{Q} (DeCarlo, 2012; Templin & Henson, 2006), estimating \mathbf{Q} for the deterministic-input, noisy-and-gate (DINA) model (Y. Chen, Culpepper, Chen, & Douglas, 2016; Chung, 2014; J. Liu, Xu, & Ying, 2012, 2013; Xiang, 2013), and estimating \mathbf{Q} for more general CDMs (Y. Chen, Liu, Xu, & Ying, 2015; Xu & Shang, 2017). Estimation of the DINA \mathbf{Q} was an important step for broadening the application of CDMs. However, the DINA model is one of the more restrictive CDMs. That is, if K denotes the number of binary attributes, the DINA model assumes that responses for members of the 2^K latent classes are described by only two parameters for each item, which may be too restrictive in some cases. Consequently, there is an opportunity to develop exploratory methods for estimating \mathbf{Q} for more flexible CDMs. One alternative to the DINA is the reduced reparameterized unified model (rRUM). The rRUM has at most $K + 1$ item parameters to describe response probabilities for the 2^K classes and is therefore a more flexible model than the DINA. It should be noted that the rRUM is a special case of several more general CDMs (de la Torre, 2011; Henson, Templin, & Willse, 2009; von Davier, 2008; Xu, 2017) that offer even greater flexibility by estimating at most 2^K parameters per item. We consider the rRUM in this article given the model is more flexible than the DINA and estimation of the rRUM \mathbf{Q} is an important development for estimating \mathbf{Q} for more general CDMs.

We develop a Bayesian method for estimating the rRUM \mathbf{Q} , which we refer to as the *exploratory* rRUM (i.e., E-rRUM). The developed E-rRUM explicitly addresses three issues associated with estimating \mathbf{Q} . First, the item response function (IRF) for the rRUM presents a unique identifiability issue that must be addressed in order to accurately estimate \mathbf{Q} . In the next section, we show that a confound arises when attempting to jointly estimate elements of \mathbf{Q} and the corresponding rRUM item parameters. Consequently, accurate recovery of the rRUM \mathbf{Q} requires directly addressing the confound between model parameters. We employ a spike-slab prior for the item parameters to address the confounding with elements of \mathbf{Q} and to select which attributes are needed for which items.

Second, Y. Chen, Culpepper, Chen, and Douglas (2016) report Monte Carlo evidence that the accuracy of Bayesian estimation of the DINA \mathbf{Q} matrix declines as the sample size N and number of attributes K increase, which is undesirable given that more flexible CDMs such as the rRUM may require more data for accurate

parameter recovery. We use Bayesian model selection procedures to more accurately estimate \mathbf{Q} using Markov chain Monte Carlo (MCMC). In fact, our Monte Carlo simulation results suggest that recovery of \mathbf{Q} improves with larger N .

Third, in our preliminary investigations, we found that estimating \mathbf{Q} is more difficult whenever the associations among the attributes satisfy a more parsimonious structure. Accordingly, there may be some instances where a parsimonious model for attributes provides better fit than an unstructured model (e.g., see Culpepper & Hudson, 2017). We incorporate a higher order model (de la Torre & Douglas, 2004; Maris, 1999) for attributes to provide a more parsimonious model for the 2^K latent classes and offer a new Gibbs sampling algorithm to estimate model parameters.

The remainder of this article includes five sections. The first section discusses model identifiability issues associated with estimating the rRUM \mathbf{Q} . In the second section, we introduce a Bayesian formulation for estimating the E-rRUM model parameters. The third section reports results from a Monte Carlo simulation study and provides evidence the developed procedures outperform prior strategies (e.g., see Y. Chen, Liu, et al., 2015; Chung, 2014) for estimating the rRUM \mathbf{Q} . The fourth section presents an application of the E-rRUM to the Examination for the Certificate of Proficiency in English (ECPE) data (Templin & Hoffman, 2013) and compares the model fit of the E-rRUM with the rRUM using an expert-specified \mathbf{Q} and a two-parameter item response theory (IRT) model. The final section discusses the implications of this study and provides concluding remarks.

The rRUM

Let Y_{ij} be the observed binary response for individual i ($1 \leq i \leq N$) to item j ($1 \leq j \leq J$). Let $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{iK})'$ where $\alpha_i \in \{0, 1\}^K$ and α_{ik} is the latent binary attribute for individual i on attribute k ($1 \leq k \leq K$). Furthermore, let $\mathbf{q}_j' = (q_{j1}, \dots, q_{jK})$ be the j th row of \mathbf{Q} such that $q_{jk} = 1$ if attribute k is required for item j and 0 otherwise. The IRF for the rRUM is

$$P(Y_{ij} = 1 | \boldsymbol{\alpha}_i, \pi_j^*, \mathbf{r}_j^*, \mathbf{q}_j) = \pi_j^* \prod_{k=1}^K r_{jk}^* (1 - \alpha_{ik})^{q_{jk}}, \quad (1)$$

where π_j^* is the probability, $Y_{ij} = 1$ if individual i has all the required attributes, $\mathbf{r}_j^* = (r_{j1}^*, \dots, r_{jK}^*)'$, and $0 \leq r_{jk}^* \leq 1$ is a penalty parameter for missing a required attribute k .

One observation from Equation 1 is that there is a confound between q_{jk} and r_{jk}^* . Specifically, it is impossible to distinguish between rRUM IRFs, where $q_{jk} = 0$ versus $q_{jk} = 1$ and $r_{jk}^* = 1$ given that

$$P(Y_{ij} | \boldsymbol{\alpha}_i, \mathbf{r}_j^*, \pi_j^*, \mathbf{q}_j, r_{jk}^* = 1, q_{jk} = 1) = P(Y_{ij} | \boldsymbol{\alpha}_i, \mathbf{r}_j^*, \pi_j^*, \mathbf{q}_j, q_{jk} = 0). \quad (2)$$

Clearly, the estimation of q_{jk} is confounded with r_{jk}^* . Furthermore, researchers can expect empirical identifiability issues whenever $r_{jk}^* \approx 1$. In fact, during our preliminary investigations estimating the rRUM \mathbf{Q} via MCMC, we generally found that a true $q_{jk} = 0$ was estimated as $q_{jk} = 1$ with a corresponding $r_{jk}^* \approx 1$. Therefore, successful estimation of the rRUM \mathbf{Q} must directly address the confound between q_{jk} and r_{jk}^* .

In the next section, we present a strategy for estimating the E-rRUM. That is, we fix the elements of \mathbf{Q} equal to 1 (i.e., $q_{jk} = 1$ for all j and k) and perform model selection to determine whether the associated item parameter is $r_{jk}^* = 1$ (i.e., the attribute is not needed for the item) or $0 \leq r_{jk}^* < 1$ (i.e., the attribute is needed for the item).

Bayesian Estimation of E-rRUM

Overview Bayesian Formulation

This subsection discusses the following three components of the E-rRUM, which includes models for the (1) observed Y_{ij} , (2) latent structure for α_i , and (3) item parameters $\boldsymbol{\pi}^* = (\pi_1^*, \dots, \pi_J^*)$ and \mathbf{r}^* .

Model for observed Y_{ij} . The E-rRUM employs a data augmentation strategy to model the rRUM IRF (e.g., see Culpepper & Hudson, 2017). The rRUM is a generalized noisy input, deterministic “AND” gate model, which assumes that observed responses are deterministically related to a collection of latent responses according to an “AND” logic gate, and the latent responses are stochastically related to the underlying attributes. Accordingly, let X_{ijk} be a latent response that equals 1 if individual i correctly applies attribute k to item j and 0 otherwise, and let $\mathbf{X}_{ij\bullet} = (X_{ij1}, \dots, X_{ijK})'$ be a random vector of latent responses for individual i on item j and let $X_{ij\bullet}$ be a realized value. We use the conjunctive condensation rule (i.e., an “AND” logic gate) to deterministically relate the observed Y_{ij} to the latent responses,

$$Y_{ij} = \prod_{k=1}^K X_{ijk}. \quad (3)$$

Unlike Culpepper and Hudson (2017), the conjunctive condensation rule in Equation 3 is not a function of the \mathbf{Q} matrix. Instead, we fix $q_{jk} = 1$ in Equation 3 to address the confound between r_{jk}^* and q_{jk} . Consequently, as discussed below, under our formulation, “inactive” latent responses are characterized by $X_{ijk} = 1$.

We model each latent response as conditionally independent given attributes and augmented item parameters,

$$X_{ijk} | \alpha_{ik}, \beta_{0jk}, \beta_{1jk} \sim \text{Bernoulli}[\Phi(\beta_{0jk} + \beta_{1jk}\alpha_{ik})], \quad (4)$$

where $\Phi(\cdot)$ is the normal cumulative distribution function, β_{0jk} is an intercept, and β_{1jk} is the impact of possessing the attribute on the probability of a correct latent response. Under this formulation, the probability of “guessing” latent response k on item j is $\Phi(\beta_{0jk})$, and the probability of slipping is $1 - \Phi(\beta_{0jk} + \beta_{1jk})$.

We use a probit data augmentation strategy to model X_{ijk} (e.g., see Albert, 1992). That is, we define a deterministic relationship between X_{ijk} and a latent augmented variable, X_{ijk}^* , such that $X_{ijk} = \mathcal{I}(X_{ijk}^* > 0)$. The prior for the continuous latent response augmented data is $X_{ijk}^* | \alpha_{ik}, \beta_{0jk}, \beta_{1jk} \sim \mathcal{N}(\beta_{0jk} + \beta_{1jk} \alpha_{ik}, 1)$.

Model for latent structure of α_i . We consider two different models for the latent attributes, α_i . First, we consider an unstructured model for α_i (e.g., see Culpepper, 2015; Culpepper & Hudson, 2017) as presented below in Equations 5 and 6:

$$p(\alpha_i | \pi) \propto \prod_{c=0}^{2^K-1} \pi_c^{\mathcal{I}(\alpha_i' \mathbf{v} = c)}, \quad \mathbf{v} = (2^{K-1}, \dots, 2, 1)', \quad (5)$$

$$\boldsymbol{\pi} = (\pi_0, \pi_1, \dots, \pi_{2^K-1})' \sim \text{Dirichlet}(\mathbf{d}_0), \quad \mathbf{d}_0 = (d_{00}, \dots, d_{0,2^K-1}), \quad (6)$$

where $\mathcal{I}(\cdot)$ is the indicator function, $\pi_c = P(\alpha_i' \mathbf{v} = c)$, and the K -vector \mathbf{v} is used to create a bijection between the 2^K classes and the integers between 0 and $2^K - 1$ (Chung, 2014; von Davier, 2014a). Prior information about class probabilities can be encoded in \mathbf{d}_0 . We use a uniform prior for π with $\mathbf{d}_0 = \mathbf{1}_{2^K}$.

In practice, the latent class probabilities may satisfy some underlying structure that is better described by a more parsimonious model. At least two alternatives have been proposed in the literature. For instance, Henson, Templin, and Willse (2009) and Templin, Henson, Templin, and Roussos (2008) describe a multivariate probit model underlying α_i , and de la Torre and Douglas (2004) and Maris (1999) assume the binary attributes are independent when conditioned upon higher order factors. We consider the higher order factor model as an alternative to the unstructured model in Equations 5 and 6 where the elements of α_i are assumed independent, given a p -vector of higher order factors, $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{ip})'$,

$$p(\alpha_i | \boldsymbol{\xi}_i, \boldsymbol{\tau}, \boldsymbol{\Lambda}) = \prod_{k=1}^K p(\alpha_{ik} | \boldsymbol{\xi}_i, \boldsymbol{\tau}, \boldsymbol{\Lambda}), \quad (7)$$

where $\boldsymbol{\tau} = (\tau_1, \dots, \tau_K)'$ is a vector of thresholds and $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_K)'$ is a $K \times p$ loading matrix. There are many options for the parametric form for the model relating α_{ik} to $\boldsymbol{\xi}_i$. We use a probit model to relate latent factors to attributes, that is, $p(\alpha_{ik} = 1 | \boldsymbol{\xi}_i, \boldsymbol{\tau}_k, \boldsymbol{\Lambda}_k) = \Phi(\boldsymbol{\Lambda}_k' \boldsymbol{\xi}_i - \tau_k)$.

Our formulation for the higher order probit model follows strategies from Bayesian IRT models (Albert, 1992; Béguin & Glas, 2001; Culpepper, 2016).

That is, we introduce a normally distributed augmented variable α_{ik}^* and specify a deterministic relationship between α_{ik} and α_{ik}^* . Let $\alpha_i^* = (\alpha_{i1}^*, \dots, \alpha_{iK}^*)'$ be a K -vector of augmented variables for individual i . Similar to de la Torre and Douglas (2004) and Templin et al. (2008), we consider a single higher-order factor ξ_i (i.e., $p = 1$) for individual i and let $\xi = (\xi_1, \dots, \xi_N)'$ (note that the formulation below can be easily extended to the $p > 1$ case). Additionally, let $\lambda = (\lambda_1, \dots, \lambda_K)'$ be a K -vector of loadings for the single factor. The Bayesian formulation for the higher order probit model with a single factor follows:

$$\alpha_{ik} = \mathcal{I}(\alpha_{ik}^* > 0), \quad (8)$$

$$\alpha^* = (\alpha_1^*, \dots, \alpha_N^*)' | \xi, \lambda, \tau \sim \mathcal{MN}_{N \times K}(\xi \lambda' - \mathbf{1}_N \tau', \mathbf{I}_N, \mathbf{I}_K), \quad (9)$$

$$\xi \sim \mathcal{N}_N(\mathbf{0}_N, \mathbf{I}_N), \quad (10)$$

$$\tau \sim \mathcal{N}_K(\mathbf{0}_K, \sigma_\tau^2 \mathbf{I}_K), \quad (11)$$

$$\lambda \sim \mathcal{N}_K(\mathbf{0}_K, \sigma_\lambda^2 \mathbf{I}_K) \prod_{k=1}^K \mathcal{I}(\lambda_k > 0). \quad (12)$$

The deterministic relationship between α_{ik} and α_{ik}^* in Equation 8 and the matrix normal distribution prior for the $N \times K$ matrix α^* in Equation 9 together imply the probit models for the higher order single factor model. Equation 10 specifies a multivariate normal prior for ξ across individuals with a N -vector of 0s (i.e., $\mathbf{0}_N$) as the expected value and an N identity matrix as the variance–covariance matrix (i.e., the higher order factors have unit variance, and independence is assumed between individuals). Equations 11 and 12 specify multivariate normal priors for the latent thresholds τ and loadings λ with means $\mathbf{0}_K$ and variance–covariance matrices $\sigma_\tau^2 \mathbf{I}_K$ and $\sigma_\lambda^2 \mathbf{I}_K$, where σ_τ^2 and σ_λ^2 are specified constants. Note the identification restriction that each $\lambda_k > 0$.

Prior for π^ and r^* .* We formulate a prior for the item parameters by using a data augmentation strategy that defines π_j^* and r_{jk}^* as functions of augmented item parameters (Hartz, 2002). Specifically, the deterministic functions relating the augmented item parameters with π_j^* and r_{jk}^* follow

$$\pi_j^* = \prod_{k=1}^K \Phi(\beta_{0jk} + \beta_{1jk}), \quad r_{jk}^* = \frac{\Phi(\beta_{0jk})}{\Phi(\beta_{0jk} + \beta_{1jk})}, \quad k = 1, \dots, K. \quad (13)$$

As noted above, we estimate the E-rRUM \mathbf{Q} by determining which attributes are active (i.e., $0 \leq r_{jk}^* < 1$) versus inactive (i.e., $r_{jk}^* = 1$) for each item. The extent to which r_{jk}^* is active is characterized by different combinations of values for β_{0jk} and β_{1jk} . For instance, Equation 13 shows that r_{jk}^* is inactive (i.e., $r_{jk}^* = 1$) whenever $\beta_{1jk} = 0$. Additionally, $r_{jk}^* \approx 1$ in cases where β_{0jk} is large and

positive. For instance, values of $\beta_{0jk} > 1.96$ correspond with guessing probabilities for the latent response X_{ijk} exceeding 0.95 and r_{jk}^* near 1. In contrast, r_{jk}^* is active when $\beta_{1jk} > 0$, and the guessing probabilities are not too large (e.g., β_{0jk} is not large and positive).

These observations imply we can specify a spike-slab prior (e.g., see George & McCulloch, 1993) for β_{0jk} and β_{1jk} to select whether r_{jk}^* is active versus inactive. Let δ_{jk} be an augmented binary variable that indicates whether attribute k is needed for item j . Specifically, if attribute k is needed for item j , then $\delta_{jk} = 1$ and $\beta_{1jk} > 0$ and β_{0jk} is not positive and large. Otherwise, if $\delta_{jk} = 0$, then β_{1jk} is near 0, and β_{0jk} is large in the positive direction. The spike-slab prior to characterize the activeness of r_{jk}^* is

$$p(\beta_{0jk}, \beta_{1jk} | \delta_{jk}) \propto \delta_{jk} f_1(\beta_{0jk}, \beta_{1jk}) + (1 - \delta_{jk}) f_0(\beta_{0jk}, \beta_{1jk}), \quad (14)$$

where f_1 denotes the “slab” and f_0 is the “spike.”

O’Hara and Sillanpää (2009) review several popular approaches for Bayesian variable selection in the statistical literature, and we follow their recommendation to use a “stochastic search variable selection” prior specification. That is, we assume the augmented item parameters are independent and define the slab as $\beta_{0jk} \sim \mathcal{N}(0, 1/v_{01})$ and $\beta_{1jk} \sim \mathcal{N}(0, 1/v_{11}) \mathcal{I}(\beta_{1jk} > 0)$ and the spike as $\beta_{0jk} \sim \mathcal{N}(0, 1/v_{00})$ and $\beta_{1jk} \sim \mathcal{N}(0, 1/v_{10}) \mathcal{I}(\beta_{1jk} > 0)$, where v_{00} , v_{01} , v_{10} , and v_{11} are constants. O’Hara and Sillanpää (2009) note that the scale constants must be tuned. We found in the Monte Carlo simulation study that using the values of $v_{00} = 0.1$, $v_{10} = 10$, $v_{01} = 4$, and $v_{11} = 2$ accurately selected the active and inactive r_{jk}^* . The values for the constants have a practical interpretation as well. We noted that active r_{jk}^* are less than 1 when β_{0jk} are not too large and β_{1jk} are positive. Therefore, setting $v_{01} = 4$ and $v_{11} = 2$ corresponds to prior variances for β_{0jk} and β_{1jk} of 0.25 and 0.5, respectively, which corresponds with the notion of an active r_{jk}^* . In contrast, $v_{00} = 0.1$ implies the variance of β_{0jk} is 10, and $v_{10} = 10$ implies the variance of β_{1jk} is 0.1, so the spike allows β_{0jk} to be large and positive and β_{1jk} to be near 0 as characterized by inactive r_{jk}^* .

The spike-slab model selection parameters δ_{jk} for $j = 1, \dots, J$ and $k = 1, \dots, K$ are conditionally independent Bernoulli random variables with probability ω ,

$$\delta_{jk} | \omega \overset{\text{ind.}}{\sim} \text{Bernoulli}(\omega), \quad (15)$$

$$\omega \sim \text{Beta}(a, b). \quad (16)$$

Furthermore, ω follows a Beta distribution with prior parameters “ a ” and “ b .” We employed a uniform prior for ω in the simulation study and application and set $a = b = 1$.

The spike–slab approach in Equations 14 through 16 offers at least two benefits. First, the prior directly addresses the confound when estimating q_{jk} and δ_{jk} . Specifically, the elements of \mathbf{Q} are set to 1, and δ_{jk} is used to infer which attributes are needed for which items. Second, each δ_{jk} can be directly used to perform model selection. For instance, one option for estimating q_{jk} is to use the posterior average $\bar{\delta}_{jk}$ as $\hat{q}_{jk} = \mathcal{I}(\bar{\delta}_{jk} > 0.5)$.

Approximating the Posterior Distribution

We use Gibbs sampling to estimate the E-rRUM parameters. Specifically, the algorithm sequentially updates the (1) augmented latent responses $X_{ij\bullet}$, (2) latent attributes α_i under either the unstructured or higher order factor model, (3) rRUM item parameters, and (4) spike–slab parameters for performing model selection. Additional details regarding each step are available in Appendix A of the online version of the journal.

Monte Carlo Simulation Study

Overview

We report results from a Monte Carlo simulation study in this section to assess the relative performance of the developed Bayesian algorithms with a Bayesian method that employs a Metropolis–Hastings (MH) sampler (Chung, 2014) and a L_1 regularization estimator (Y. Chen, Liu, et al., 2015). Chung (2014) used MH sampling to update the item parameters π^* and r^* and a Gibbs sampler for updating rows of \mathbf{Q} by sampling q_j from the $2^K - 1$ nonzero configurations. The L_1 estimator penalizes the likelihood function for model complexity and simulation evidence suggests the approach is accurate for recovering the DINA model \mathbf{Q} . It is important to note that the MH sampler does not explicitly address the confound between \mathbf{Q} and r^* , so we expect the E-rRUM to be more accurate.

In the simulation study, we considered two samples sizes (i.e., $N = 1,000$ and $2,000$), four levels of attribute dependence (i.e., one condition with uniformly distributed attributes and three conditions with dependence generated from the multivariate normal threshold model with a latent correlation of $\rho = .05, .15$, and $.25$), and two values for the total number of attributes (i.e., $K = 3$ and 4). For the conditions with uniformly distributed attributes, attributes were uniformly sampled from the 2^K classes. In contrast, dependent attributes were generated from the latent multivariate normal probit model (e.g., see Chiu, Douglas, & Li, 2009) with thresholds for attribute k defined as $k/(K + 1)$, a mean vector of zero, and a constant correlation among the attributes of ρ . Following Chiu and Köhn (2016), the item parameters were defined as $\pi_j^* = 0.9$ and $r_{jk}^* = 0.6$. Furthermore, $J = 30$ for all conditions, and the true \mathbf{Q} was sampled each replication by

first randomly placing three \mathbf{I}_K matrices into the rows of \mathbf{Q} and then filling the remaining rows by sampling uniformly from the $2^K - 1$ nonzero arrangements.

We compared our unstructured and higher order single factor models discussed in the previous section with parameter estimates using the MH sampler and L_1 estimator. For our variable selection method, we fixed $\nu_{00} = 0.1$, $\nu_{10} = 10$, $\nu_{01} = 4$, and $\nu_{11} = 2$. Note the single, higher order factor model was estimated with the constraint that loadings were constant across k such that $\lambda_k = \lambda$, whereas the elements of $\boldsymbol{\tau}$ were freely estimated. The L_1 estimator requires a specified value for the lasso penalty parameter. We considered a grid of values for the penalty parameter from 0.01 to 0.15 in increments of 0.005, and we report the most accurate set of results for each replication.

The outcomes of interest were the accuracy of the estimated \mathbf{Q} across the four methods. We computed accuracy both matrix-wise and entry-wise across the replications. That is, the matrix-wise accuracy rate was computed as

$$\frac{1}{100} \sum_{r=1}^{100} \mathcal{I}(\mathbf{Q} = \hat{\mathbf{Q}}_r), \quad (17)$$

where $\hat{\mathbf{Q}}_r$ is the estimated \mathbf{Q} for replication r . The entry-wise accuracy rate was defined as

$$\frac{1}{100} \sum_{r=1}^{100} \frac{1}{JK} \sum_{j=1}^J \sum_{k=1}^K \mathcal{I}(q_{jk} = \hat{q}_{jk,r}), \quad (18)$$

where $\hat{q}_{jk,r}$ is the estimated q_{jk} for replication r .

For the MH sampler, we followed Chung (2014) and computed $\hat{\mathbf{Q}}$ using the entry-wise mode. For the E-rRUM, we computed \hat{q}_{jk} using the entry-wise mode for δ_{jk} (i.e., $\hat{q}_{jk} = \mathcal{I}(\bar{\delta}_{jk} > 0.5)$). Note the L_1 estimator for $\hat{\mathbf{Q}}$ is implied by the pattern of nonzero coefficients in the transformed rRUM IRF (e.g., see Y. Chen, Liu, et al., 2015, p. 856).

We executed 100 replications for each combination of parameter values and used chain lengths of 80,000 and a burnin of 20,000 iterations for both Bayesian methods. The Bayesian methods were programmed using Rcpp (Eddelbuettel et al., 2011). The run times for the $N = 2,000$ and $K = 4$ conditions on a cluster with 2.4 GHz processors was 2 hours, 16 hours, and 20 minutes for the E-rRUM, MH sampler, and L_1 estimator, respectively.

Monte Carlo Results

Table 1 summarizes the results from the Monte Carlo simulation study regarding the accuracy of \mathbf{Q} estimation for the methods developed in this article (i.e., the higher order and unstructured models for $\boldsymbol{\alpha}_i$), the MH sampler, and the L_1 regularization estimator. We note three features about the Monte Carlo results.

TABLE 1.

Summary of Monte Carlo Simulation Study Accuracy of \mathbf{Q} Estimation for the Higher Order Single Factor Model (HO), Unstructured Model (Uns.), Metropolis–Hastings Sampler (MH), and L_1 Regularization for Values of K , N , and ρ

N	ρ^a	K	$\mathbf{Q} = \hat{\mathbf{Q}}$				Entry-Wise Average			
			HO	Uns.	MH	L_1	HO	Uns.	MH	L_1
1,000	.00	3	70	70	0	2	98.6	98.9	76.6	86.2
2,000	.00	3	77	81	0	3	99.3	99.3	69.9	85.3
1,000	.05	3	53	58	0	0	97.5	98.6	73.2	84.8
2,000	.05	3	71	70	0	0	99.2	98.5	72.1	84.6
1,000	.15	3	52	53	0	0	97.1	97.8	76.6	86.1
2,000	.15	3	69	67	0	2	98.8	97.3	73.2	86.1
1,000	.25	3	38	43	0	0	97.6	96.6	76.6	86.1
2,000	.25	3	72	57	0	2	99.1	96.2	75.4	89.5
1,000	.00	4	49	53	0	0	98.1	98.2	67.5	80.1
2,000	.00	4	68	66	0	0	99.0	99.0	65.8	79.7
1,000	.05	4	13	10	0	0	96.6	93.9	68.3	77.7
2,000	.05	4	52	40	0	0	98.6	97.6	65.0	78.2
1,000	.15	4	23	9	0	0	96.8	94.4	68.3	79.6
2,000	.15	4	45	32	0	0	98.3	97.1	68.3	80.7
1,000	.25	4	15	10	0	0	96.3	93.2	71.7	79.8
2,000	.25	4	30	38	0	0	97.3	96.9	64.2	81.0

Note. Results are based upon 100 replications. MH = Metropolis–Hastings sampling approach of Chung (2014).

^a $\rho = 0$ denotes attributes were uniformly sampled from $\{0, 1\}^K$.

First, the Monte Carlo results support the use of the developed methods in this article to estimate the E-rRUM \mathbf{Q} matrix. That is, the higher order and unstructured models had better matrix-wise and entry-wise recovery rates than both the MH sampler and L_1 methods.

Second, the results in Table 1 provide evidence that matrix-wise recovery of \mathbf{Q} is more challenging for the rRUM than for the DINA model. For example, Y. Chen, Liu, Xu, and Ying (2015) and Y. Chen, Culpepper et al. (2016) report matrix-wise recovery rates in the 80s and 90s for the DINA for the conditions studied in this article. The Monte Carlo results suggest that the additional complexity of the rRUM affects the recovery of \mathbf{Q} . It is important to note that while the matrix-wise recovery rates are smaller for the rRUM than the DINA that the entry-wise recovery rates exceed 95% for the higher order model across the simulated conditions. Additionally, the matrix-wise recovery rates of \mathbf{Q} improve as N increased.

Third, the relative performance of our unstructured and higher order models varied by condition. Namely, the results in Table 1 suggest that the higher order

model tended to have higher matrix-wise recovery rates for larger N and ρ . Furthermore, the entry-wise recovery rates for the E-rRUM models were comparable for the $K = 3$ and 4 conditions.

Application to the “ECPE” Data

We analyze the ECPE data set (e.g., see Templin & Hoffman, 2013), which is available in the “cdm” R package version 6.4-23 (Robitzsch, Kiefer, George, & Uenlue, 2015). The ECPE data include item responses from 2,922 examinees at the University of Michigan as a test of advanced English skills. The ECPE data were originally analyzed with the rRUM by Henson and Templin (2007b) using an expert-specified \mathbf{Q} matrix, which consists of three attributes: (1) morphosyntactic rules, (2) cohesive rules, and (3) lexical rules. Subsequent studies analyzed the ECPE using clustering algorithms (Chiu et al., 2009) and confirmatory CDMs such as the DINA (Y. Liu, Douglas, & Henson, 2009), the rRUM (Chiu & Köhn, 2016; Henson & Templin, 2007b; Y. Liu et al., 2009), and more general CDMs (Templin & Bradshaw, 2014; Templin & Hoffman, 2013; von Davier, 2014b).

Recent research with the ECPE data set offered competing conclusions regarding the underlying structure. Specifically, Templin and Bradshaw (2014) and von Davier (2014b) analyzed the ECPE data set with a general CDM using the expert specified \mathbf{Q} and found evidence that the probability of membership in some classes was near 0 (i.e., that several elements of π are near 0). Researchers offered two competing inferences based on this finding. First, one inference was that the attributes are arranged into a linear hierarchy where some attributes must be mastered before others (Templin & Bradshaw, 2014). Second, an alternative interpretation is that a single, ordered skill or a continuous unidimensional trait might underlie performance on ECPE items (von Davier, 2014b; von Davier & Haberman, 2014).

We apply the unstructured E-rRUM to the ECPE data to estimate \mathbf{Q} and provide new evidence about the underlying structure. Specifically, we fit the unstructured E-rRUM with $K = 2-7$ and compare the exploratory models with the confirmatory rRUM that uses an expert-specified \mathbf{Q} matrix from Templin and Hoffman (2013) (e.g., see Table 2) using the Bayesian estimation algorithm of Culpepper and Hudson (2017). Additionally, von Davier (2014b) found evidence that a two-parameter IRT model had a lower Akaike information criterion than the rRUM with an expert-specified \mathbf{Q} , and we accordingly compare model fit of a two-parameter normal ogive (2PNO; Albert, 1992) model with the E-rRUM. We used chain of lengths of 80,000 and discard the first 20,000 iterations as burn-in. The E-rRUMs were all estimated using values for ν_{00} , ν_{01} , ν_{10} , and ν_{11} as specified for the simulation study. In the remainder of this section, we first discuss model fit and then present results from the best fitting model.

TABLE 2.
Summary of ECPE Data Unstructured E-rRUM Model Parameter Estimates for $K = 5$

Item	Expert \mathbf{Q}			E-rRUM $\widehat{\mathbf{Q}}$					$\bar{\delta}_{j1}$	$\bar{\delta}_{j2}$	$\bar{\delta}_{j3}$	$\bar{\delta}_{j4}$	$\bar{\delta}_{j5}$	\hat{r}_{j1}^*	\hat{r}_{j2}^*	\hat{r}_{j3}^*	\hat{r}_{j4}^*	\hat{r}_{j5}^*	$\hat{\pi}_j^*$
	q_{j1}	q_{j2}	q_{j3}	\hat{q}_{j1}	\hat{q}_{j2}	\hat{q}_{j3}	\hat{q}_{j4}	\hat{q}_{j5}											
1	1	1	0	1	0	1	0	0	1.00	0.01	0.90	0.03	0.45	0.82	1.00	0.84	0.99	0.94	0.93
2	0	1	0	1	0	1	0	0	1.00	0.01	0.80	0.18	0.02	0.85	1.00	0.87	0.97	0.99	0.92
3	1	0	1	1	0	1	0	1	1.00	0.19	0.99	0.37	0.90	0.79	0.97	0.71	0.91	0.81	0.81
4	0	0	1	1	1	0	1	1	1.00	0.58	0.29	0.71	0.62	0.66	0.86	0.95	0.82	0.88	0.89
5	0	0	1	1	0	0	0	0	1.00	0.19	0.00	0.17	0.07	0.69	0.97	1.00	0.97	0.99	0.96
6	0	0	1	1	0	0	0	1	1.00	0.13	0.25	0.10	0.83	0.77	0.98	0.97	0.98	0.89	0.96
7	1	0	1	1	0	1	0	1	1.00	0.17	1.00	0.02	0.86	0.56	0.96	0.77	0.99	0.85	0.95
8	0	1	0	1	0	1	0	0	0.99	0.00	0.51	0.46	0.00	0.89	1.00	0.94	0.92	1.00	0.97
9	0	0	1	1	1	0	1	1	0.94	0.50	0.15	0.52	0.84	0.80	0.89	0.97	0.86	0.83	0.87
10	1	0	0	1	0	1	0	1	1.00	0.02	0.99	0.19	1.00	0.80	1.00	0.76	0.96	0.74	0.90
11	1	0	1	1	0	0	0	1	1.00	0.16	0.47	0.30	1.00	0.70	0.97	0.93	0.94	0.71	0.93
12	1	0	1	1	1	1	0	1	1.00	0.51	0.81	0.03	1.00	0.51	0.83	0.78	0.99	0.34	0.77
13	1	0	0	1	0	1	0	0	1.00	0.20	0.91	0.28	0.43	0.73	0.97	0.85	0.94	0.94	0.92
14	1	0	0	1	0	1	0	1	0.99	0.04	0.78	0.43	0.95	0.82	0.99	0.81	0.90	0.82	0.84
15	0	0	1	1	0	1	0	0	1.00	0.08	0.62	0.00	0.00	0.64	0.99	0.94	1.00	1.00	0.97
16	1	0	1	1	0	1	0	1	1.00	0.19	1.00	0.00	0.80	0.60	0.96	0.76	1.00	0.86	0.93
17	0	1	1	1	0	1	0	0	1.00	0.00	0.56	0.00	0.01	0.80	1.00	0.93	1.00	1.00	0.95
18	0	0	1	1	0	0	0	0	1.00	0.18	0.24	0.44	0.00	0.73	0.97	0.96	0.93	1.00	0.93
19	0	0	1	1	1	0	0	0	1.00	1.00	0.28	0.27	0.25	0.73	0.56	0.96	0.94	0.95	0.91
20	1	0	1	1	1	1	0	1	1.00	0.86	1.00	0.26	1.00	0.62	0.69	0.69	0.92	0.53	0.82
21	1	0	1	1	0	1	0	1	1.00	0.20	0.55	0.31	0.98	0.67	0.95	0.91	0.94	0.84	0.93

(continued)

TABLE 2. (continued)

Item	Expert \mathcal{Q}					E-rRUM $\widehat{\mathcal{Q}}$										\hat{r}_{j5}^*	\hat{r}_{j4}^*	\hat{r}_{j3}^*	\hat{r}_{j2}^*	\hat{r}_{j1}^*	$\bar{\delta}_{j5}$	$\bar{\delta}_{j4}$	$\bar{\delta}_{j3}$	$\bar{\delta}_{j2}$	$\bar{\delta}_{j1}$
	q_{j1}	q_{j2}	q_{j3}	\hat{q}_{j1}	\hat{q}_{j2}	\hat{q}_{j3}	\hat{q}_{j4}	\hat{q}_{j5}																	
22	0	0	1	1	1	0	0	1	1.00	0.66	0.35	0.04	0.99	0.42	0.81	0.94	0.99	0.70	0.87						
23	0	1	0	1	0	1	1	1	1.00	0.01	0.60	0.60	0.52	0.77	1.00	0.91	0.90	0.93	0.95						
24	0	1	0	1	0	1	1	1	1.00	0.12	0.56	0.53	0.96	0.68	0.98	0.85	0.80	0.79	0.73						
25	1	0	0	1	0	1	0	0	0.92	0.42	0.98	0.16	0.40	0.89	0.92	0.72	0.97	0.92	0.81						
26	0	0	1	1	0	1	0	0	1.00	0.16	0.64	0.06	0.44	0.68	0.98	0.90	0.99	0.93	0.82						
27	1	0	0	1	1	1	1	1	0.99	0.74	0.97	0.51	0.97	0.79	0.76	0.71	0.77	0.72	0.75						
28	0	0	1	1	1	0	0	0	1.00	0.99	0.01	0.00	0.23	0.74	0.74	1.00	1.00	0.97	0.94						

Note. The expert \mathbf{Q} is from Henson and Templin (2007b). $\hat{q}_{jk} = \mathcal{I}(\bar{\delta}_{jk} > 0.5)$. $\bar{\delta}_{jk}$ and r_{jk}^* for $k = 1, 2, 3, 4, 5$ and $\hat{\pi}_j^*$ are posterior averages. ECPE = Examination for the Certificate of Proficiency in English; E-rRUM = exploratory reduced reparameterized unified model.

Model Fit

We first compare the relative fit of the models using the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). Specifically, we computed the marginal DIC (i.e., we marginalize the likelihood over attributes for the diagnostic models and integrate over the latent traits for the 2PNO) rather than a conditional DIC for missing data models (e.g., see Celeux, Forbes, Robert, & Titterton, 2006). The smallest DIC value of 85,058 corresponded to the E-rRUM with $K = 5$. In contrast, the DICs for the expert \mathbf{Q} rRUM and the 2PNO were 111,143 and 85,203, respectively, and the DICs for the E-rRUM with $K = 2, 3, 4, 6$, and 7 were 85,258, 85,202, 85,124, 85,074, and 85,079.

We also evaluated relative model fit using posterior predictive checks. In particular, the \mathbf{Q} matrix specifies the underlying structure, so one way to evaluate relative model fit is to compare how well the models reproduce the item means and observed relationships among items. That is, an inadequate model will be less likely to predict item means and reproduce the associations among the observed variables. We therefore assess model fit by computing posterior predictive probabilities (PPPs) of the item means and odds ratios for each pair of items (e.g., see W.-H. Chen & Thissen, 1997; Sinharay, Johnson, & Stern, 2006). We follow Sinharay, Johnson, and Stern (2006) and consider PPPs smaller than 0.05 or greater than 0.95 to be extreme and evidence of misfit.

We found evidence that all of the models effectively reproduced the observed item means, so we do not report those results here. We computed odds-ratio PPPs by (1) simulating observed responses $\mathbf{Y}^{(r)}$ using model parameters from iteration r of the MCMC sampler; (2) computing the odds ratio for each pair of items at iteration r as $OR^{(r)} = n_{11}^{(r)}n_{00}^{(r)} / (n_{10}^{(r)}n_{01}^{(r)})$, where $n_{11}^{(r)}$ is the frequency of ones on both variables at iteration r , $n_{10}^{(r)}$ is the frequency of ones on the first item and zeros on the second at iteration r , etc.; and (3) computing PPPs for each item pair as the proportion of generated $OR^{(r)}$'s that exceeded elements of the observed odds ratios. We found evidence the E-rRUM better predicted relationships among the ECPE items than the rRUM with an expert-specified \mathbf{Q} and the 2PNO. In fact, 28.3% of the expert \mathbf{Q} rRUM pairwise odd ratios were considered too extreme (i.e., the confirmatory rRUM did a poor job of describing pairwise item odds ratios), and 10.8% of the 2PNO PPPs were considered extreme. In contrast, the percentage of out-of-range odds-ratio PPPs for the E-rRUM were 11.1%, 10.3%, 7.4%, 4.8%, 5.8%, and 5.8% for $K = 2, 3, 4, 5, 6$, and 7 . The DICs and odds-ratio PPPs support the E-rRUM with $K = 5$. In the next subsection, we summarize the results of the empirical application of the E-rRUM with $K = 5$.

Empirical Results for $K = 5$

The results in the previous subsection support the presence of five binary attributes rather than a continuous latent trait. In this subsection, we report results for the E-rRUM with $K = 5$ and discuss the extent to which there is evidence for an attribute hierarchy. Table 2 reports the E-rRUM parameter estimates (i.e., \hat{Q} , $\hat{\delta}$, \hat{r}^* , and $\hat{\pi}^*$) along with the previously employed expert Q . Note that \hat{Q} is defined using the posterior element-wise mode for δ (i.e., $\hat{q}_{jk} = \mathcal{I}(\hat{\delta}_{jk} > 0.5)$). One immediate observation from Table 2 is that the expert Q is not captured by \hat{Q} , which may be expected given the differences in model fit discussed in the previous subsection.

The results in Table 2 offer evidence about the extent to which there is an attribute hierarchy. Specifically, Table 2 shows that the estimated Q implies the first attribute is needed for all items. The posterior averages of δ_{jk} in Table 2 quantify uncertainty regarding the corresponding \hat{q}_{jk} , and the estimates suggest that the posterior probability the first attribute is required exceeds 0.90 for all items. One conclusion is that students must possess the first attribute to have the greatest probability of successfully answering the items. In fact, the penalty for not possessing the first attribute is as low as $r_{7,1}^* = 0.56$ and $r_{22,1}^* = 0.42$ for items 7 and 22, and the average penalty for not possessing Attribute 1 equaled 0.72.

Additional evidence to support an attribute hierarchy is found in the estimated attribute structure. Table 3 reports the estimated ECPE latent class proportions (i.e., $\hat{\pi}$) for the E-rRUM with $K = 5$. Our results in Table 3 agree with prior research given that the probability of membership in several classes are near 0. Specifically, the estimated π suggests there is a smaller probability of being classified into an attribute class that is missing the first attribute. In fact, only 9.8% of the sample was assigned to one of the 14 classes that possesses at least one attribute other than α_1 . Table 3 shows that seven of the 32 possible latent classes had membership proportions exceeding 5%. Specifically, the structural probabilities for the seven largest latent classes were .062 for (00000), .060 for (10000), .060 for (11000), .075 for (11001), .081 for (11010), .086 for (11011), and .242 for (11111).

Templin and Bradshaw (2014) interpreted the class probabilities and concluded there was evidence of a linear attribute hierarchy. The results in Table 3 may instead provide evidence of a more complex attribute hierarchy (e.g., see Köhn & Chiu, 2018). That is, using 5% as a threshold to infer which classes are nonzero implies that Attribute 2 requires Attribute 1 is mastered, Attributes 4 and 5 require Attribute 2, and Attribute 3 requires Attributes 4 and 5 are mastered. Rather than a linear hierarchy, the estimated class probabilities provide evidence of a partially ordered attribute hierarchy, given Attributes 4 and 5 are not ordered.

TABLE 3.

ECPE Data Estimated Latent Class Proportions $\hat{\pi}$ for Unstructured E-rRUM With $K = 5$

α	$\hat{\pi}$	α	$\hat{\pi}$
00000	.062	10000	.060
00001	.005	10001	.011
00010	.012	10010	.041
00011	.006	10011	.013
00100	.008	10100	.014
00101	.005	10101	.012
00110	.010	10110	.029
00111	.007	10111	.011
01000	.010	11000	.060
01001	.005	11001	.075
01010	.008	11010	.081
01011	.007	11011	.086
01100	.005	11100	.033
01101	.005	11101	.028
01110	.006	11110	.030
01111	.012	11111	.242

Note. ECPE = Examination for the Certificate of Proficiency in English; E-rRUM = exploratory reduced reparameterized unified model.

Discussion

We presented a new exploratory approach for estimating the rRUM \mathbf{Q} matrix. We conclude this article in this section with a summary of our contribution and recommendations for future research.

The E-rRUM provides a more flexible modeling framework than previous research that estimated the DINA \mathbf{Q} . Specifically, the more restrictive DINA model includes only 2-item parameters, whereas the rRUM has up to $K + 1$ parameters per item. It is important to note that the increased flexibility of the E-rRUM translated to lower matrix-wise recovery rates of \mathbf{Q} than observed in prior research for the DINA model. This finding has implications for future research on estimating \mathbf{Q} for more general CDMs. Namely, the rRUM is a special case of several general CDMs (de la Torre, 2011; Henson et al., 2009; von Davier, 2008), and it is likely that estimating \mathbf{Q} will be more challenging for these models than the rRUM. An important goal for future research is to develop procedures for estimating \mathbf{Q} for more general CDMs. Xu and Shang's (2017) results provide a theoretical foundation for such work, and future research must necessarily examine how study design features (e.g., sample size, latent attribute structure, the number of attributes) impact recovery of \mathbf{Q} for these general models.

We also offer improved Bayesian strategies for estimating the \mathbf{Q} matrix. The Monte Carlo evidence supports the use of the developed procedures over existing Bayesian methods for estimating the rRUM \mathbf{Q} . In the simulation study, the MH sampler did not in any replication or condition recover all elements of \mathbf{Q} . The difference in model performance is likely attributed to the fact that our developed Bayesian algorithms addressed the confound between \mathbf{Q} and \mathbf{r}^* . Specifically, we addressed the confound by fixing all elements of \mathbf{Q} equal to 1 and then used spike-slab priors for the augmented item parameters to determine which penalty parameters are active. The Monte Carlo results suggest the developed methods improve upon existing research. Another important finding from the Monte Carlo study is that recovery of \mathbf{Q} improved as N increased for $K = 3$ and $K = 4$.

The developed E-rRUM also offers a practical contribution for applied researchers. The E-rRUM is applicable for theory development in a manner similar to how exploratory factor analysis has been applied across the social sciences. In fact, our application of the E-rRUM to the ECPE data provides examples of the types of inferences that are available with exploratory CDMs. First, we found evidence that estimating \mathbf{Q} improved model fit in comparison to using an expert-specified \mathbf{Q} matrix and a unidimensional latent trait. Accordingly, future researchers can use the E-rRUM to evaluate the plausibility of existing cognitive theory. Second, we uncovered a different latent structure for attributes than found in previous research using an expert-specified \mathbf{Q} matrix. One implication is that estimating \mathbf{Q} has the potential to impact inferences and theory development. For example, as noted above, prior research concluded that the attributes underlying the ECPE data either satisfied a hierarchy or were ordered along a unidimensional trait. The results of the E-rRUM instead support the existence of five hierarchically structured attributes. That is, we found evidence from the estimated \mathbf{Q} that the first attribute is required for all items, and the pattern of latent class probabilities may support a partially ordered attribute hierarchy. Consequently, using the E-rRUM rather than the rRUM improved model fit and altered substantive conclusions regarding the nature of the latent attribute structure.

There are several directions for research. First, although the Monte Carlo simulation study employed a similar design as previous studies, there is an opportunity in future research to examine model parameter accuracy for additional attribute latent structures, numbers of attributes, and sample sizes. Subsequent Monte Carlo research will improve our understanding of factors that impact recovery of \mathbf{Q} . For instance, additional evidence is needed about parameter recovery and classification accuracy for cases where a subset of the items might not load on the common skills and attributes. Second, future research should consider extending the methods developed to more general CDMs. Our application of the E-rRUM to the ECPE data set assumed a conjunctive relationship, which may not be appropriate for all items. Future research should apply more general exploratory CDMs to the ECPE data to evaluate the plausibility of

the conjunctive assumption and to further evaluate the presence of an attribute hierarchy. Third, we considered a fully exploratory approach for estimating \mathbf{Q} ; however, researchers may have partial information where some values are known based upon cognitive theory. In such cases, if some q_{jk} are known, our approach can be used by updating the remaining elements of \mathbf{Q} conditioned upon the known values. Future research should also more generally consider how to incorporate expert knowledge into estimation. One approach researchers may consider is to replace our prior with one that uses expert knowledge to specify the chance each entry is 0 versus 1.

A final direction for future research relates to model identifiability for exploratory CDMs. For instance, the estimated ECPE \mathbf{Q} reported in Table 2 does not satisfy sufficient conditions that ensure the model parameters are strictly identified (i.e., that the model parameters map to distinct likelihoods). Specifically, Xu and Shang (2017) showed that one of the sufficient conditions to strictly identify CDM parameters is that \mathbf{Q} includes two identity matrices (see Appendix B, available in the online version of the journal, for a summary of Xu and Shang's [2017] results). The estimated ECPE \mathbf{Q} matrix for $K = 5$ does not include simple structure for 10 items. Consequently, it may be the case that the reported E-rRUM parameters are not strictly identified. In preliminary analyses, we fit an E-rRUM that explicitly enforced model identifiability results (see Appendix B, available in the online version of the journal, for details about the algorithm for enforcing identifiability constraints). The results from the constrained E-rRUM supported a solution with $K = 3$, and we uncovered an estimated \mathbf{Q} with dense structure that included 6 simple structure items and the remaining elements equal to 1s. One possible explanation for uncovering a dense \mathbf{Q} is that the true \mathbf{Q} may not satisfy Xu and Shang's (2017) sufficient conditions for strict identifiability, and enforcing the identifiability conditions may yield a misspecified model. Although the estimated \mathbf{Q} we report may not satisfy conditions for strict identifiability, the model parameters may be generically identified, which "... implies that the set of points for which identifiability does not hold has measure zero. In this sense, any observed data set has probability one of being drawn from a distribution with identifiable parameters" (Allman, Matias, & Rhodes, 2009, p. 3102). Satisfying generic identifiability is likely sufficient for practical applications of exploratory CDMs, given there is no guarantee the available items satisfy the more stringent strict identifiability conditions. Additional research is clearly needed to establish generic identifiability results for more general CDMs.

In conclusion, CDMs are popular for their ability to classify individuals on a fine-grained collection of attributes. The development of exploratory methods to estimate \mathbf{Q} for more general models is critical for widespread application of CDMs. We introduced an exploratory version of the rRUM and made several advances for Bayesian estimation of \mathbf{Q} . The developed methods improved upon

existing research and further broadened the use and applicability of CDMs in applied research.

Acknowledgments

We thank Jeffrey Douglas, the editor, and two anonymous reviewers for helpful comments and suggestions and Yunxiao Chen for sharing R code for estimating the DINA Q matrix using the L_1 regularization estimator that we modified to estimate the rRUM Q matrix. Any remaining errors belong to the authors.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the Spencer Foundation Grant #201700062 and by a grant from the National Science Foundation Methodology, Measurement, and Statistics program grant #1632023. Opinions reflect those of the authors and do not necessarily reflect those of the granting agency.

References

- Albert, J. (1992). Bayesian estimation of normal ogive item response curves using Gibbs sampling. *Journal of Educational and Behavioral Statistics*, 17, 251–269.
- Allman, E. S., Matias, C., & Rhodes, J. A. (2009). Identifiability of parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37, 3099–3132.
- Béguin, A. A., & Glas, C. A. (2001). MCMC estimation and some model-fit analysis of multidimensional IRT models. *Psychometrika*, 66, 541–561.
- Celeux, G., Forbes, F., Robert, C. P., & Titterton, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1, 651–673. Retrieved from <https://doi.org/10.1214/06-BA122> doi: 10.1214/06-BA122
- Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, 22, 265–289.
- Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2016). *Bayesian estimation of the DINA Q*. Paper presented at the International Meeting of the Psychometric Society, Asheville, NC.
- Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, 110, 850–866.
- Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37, 598–618.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74, 633–665.
- Chiu, C.-Y., & Köhn, H.-F. (2016). The reduced RUM as a logit model: Parameterization and constraints. *Psychometrika*, 81, 350–370.

- Chung, M. (2014). *Estimating the Q-matrix for cognitive diagnosis models in a Bayesian framework* (Unpublished doctoral dissertation). Columbia University, NY.
- Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, 40, 454–476.
- Culpepper, S. A. (2016). Revisiting the 4-parameter item response model: Bayesian estimation and application. *Psychometrika*, 81, 1142–1163.
- Culpepper, S. A., & Hudson, A. (2017). An improved strategy for Bayesian estimation of the reduced reparameterized unified model. *Applied Psychological Measurement*, 42, 99–115. doi:10.1177/0146621617707511
- DeCarlo, L. T. (2012). Recognizing uncertainty in the Q-matrix via a Bayesian extension of the DINA model. *Applied Psychological Measurement*, 36, 447–468.
- de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement*, 45, 343–362.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179–199.
- de la Torre, J., & Chiu, C.-Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, 81, 253–273.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333–353.
- Eddelbuettel, D., François, R., Allaire, J., Chambers, J., Bates, D., & Ushey, K. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40, 1–18.
- George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88, 881–889.
- Hartz, S. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Unpublished doctoral dissertation) University of Illinois at Urbana, Champaign, IL.
- Henson, R. A., & Templin, J. (2007a). *Importance of Q-matrix construction and its effects cognitive diagnosis model results*. Paper presented at Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Henson, R. A., & Templin, J. (2007b). *Large-scale language assessment using cognitive diagnosis models*. Paper presented at Annual Meeting of the National Council on Measurement in Education, Chicago, IL.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191–210.
- Köhn, H.-F., & Chiu, C.-Y. (2018). Identifiability of the latent attribute space and conditions of Q-matrix completeness for attribute hierarchy models. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (eds), *Quantitative Psychology*, (pp. 363–375). IMPS 2017. Springer Proceedings in Mathematics & Statistics, vol 233. Springer, Cham. Retrieved from https://doi.org/10.1007/978-3-319-77249-3_30.
- Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. *Applied Psychological Measurement*, 36, 548–564.
- Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli*, 19, 1790–1817.

- Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, 33, 579–598.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187–212.
- O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, 4, 85–117.
- Robitzsch, A., Kiefer, T., George, A. C., & Uenlue, A. (2015). *CDM: Cognitive diagnosis modeling* (R package version 4.6-0) [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=CDM>
- Rupp, A. A., & Templin, J. L. (2008). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68, 78–96.
- Sinharay, S., Johnson, M. S., & Stern, H. S. (2006). Posterior predictive assessment of item response theory models. *Applied Psychological Measurement*, 30, 298–321.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 583–639.
- Templin, J. L., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79, 317–339.
- Templin, J. L., & Henson, R. A. (2006). *A Bayesian method for incorporating uncertainty into Q-matrix estimation in skills assessment*. Symposium Conducted at the Meeting of the American Educational Research Association, San Francisco, CA.
- Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement*, 32, 559–574.
- Templin, J. L., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement: Issues and Practice*, 32, 37–50.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287–307.
- von Davier, M. (2014a). The DINA model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67, 49–71.
- von Davier, M. (2014b). The log-linear cognitive diagnostic model (LCDM) as a special case of the general diagnostic model (GDM). Research Report. ETS RR-14-40. ETS Research Report Series. doi:10.1002/ets2.12043
- von Davier, M., & Haberman, S. J. (2014). Hierarchical diagnostic classification models morphing into unidimensional “diagnostic” classification models: A commentary. *Psychometrika*, 79, 340–346.
- Xiang, R. (2013). *Nonlinear penalized estimation of true Q-matrix in cognitive diagnostic models* (Unpublished doctoral dissertation). Columbia University, NY.
- Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, 45, 675–707.
- Xu, G., & Shang, Z. (2017). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*. doi:10.1080/01621459.2017.1340889

Authors

STEVEN ANDREW CULPEPPER is an associate professor in the Department of Statistics at the University of Illinois at Urbana-Champaign, Illini Hall, Room 115, 725 S. Wright Street, Champaign, IL 61820, USA; email: sculpepp@illinois.edu. His research interests include large scale testing, Bayesian models and computation, restricted latent class models, and longitudinal modeling.

YINGHAN CHEN is an assistant professor in the Department of Mathematics & Statistics at the University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557, USA; email: yinghanc@unr.edu. Her research interests include statistical computing, advanced Monte Carlo methods, Bayesian analysis, and latent class models.

Manuscript received May 18, 2017

First revision received October 30, 2017

Second revision received May 24, 2018

Accepted June 18, 2018